DOT/FAA/AM-99/21
Office of Aviation Medicine
Washington, D.C. 20591

Improving Pilot/ATC Voice Communication in General Aviation

Daniel G. Morrow
University of New Hampshire
Durham, New Hampshire 03824

O. Veronika Prinzo
Civil Aeromedical Institute
Federal Aviation Administration
Oklahoma City, Oklahoma 73125

July 1999

Final Report

This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161.

U.S. Department of Transportation
Federal Aviation Administration

Technical Report Documentation Page

1. Report No. DOT/FAA/AM-99/21	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Improving Pilot/ATC Voice Communication in General Aviation		5. Report Date July 1999
		6. Performing Organization Code
7. Author(s) Morrow, D.G.[1], and Prinzo, O.V.[2]		8. Performing Organization Report No.
9. Performing Organization Name and Address [1]Daniel G. Morrow University of New Hampshire Durham, New Hampshire 03824 [2]FAA Civil Aeromedical Institute P.O. Box 25082 Oklahoma City, OK 73125		10. Work Unit No. (TRAIS)
		11. Contract or Grant No. DTFA-02-96-P-54069
12. Sponsoring Agency name and Address Office of Aviation Medicine Federal Aviation Administration 800 Independence Ave., S.W. Washington, D.C. 20591		13. Type of Report and Period Covered
		14. Sponsoring Agency Code

15. Supplemental Notes
This work was performed under Task AM-B-96-HRR-513

16. Abstract

The influence of Air Traffic Control (ATC) instruction format (grouped vs. sequential presentation) and message length on General Aviation pilot communication was investigated in a simulated flight environment using the Civil Aeromedical Institute's (CAMI's) Basic General Aviation Research Simulator (BGARS). Prior to flying the simulator each pilot was provided with familiarization training, listened to and read back ATC messages spoken in either grouped or sequential format (depending on their assigned treatment group), and completed a digit span test (a measure of short-term memory). While flying 2 missions in the simulator, 12 pilots heard recorded ATC messages that contained altitude and radio frequency information spoken in a grouped format (e.g., "descend and maintain forty-one hundred"), and 12 heard the same instructions spoken sequentially (e.g., "descend and maintain four thousand one hundred"). The amount of information in a message varied from 2 to 5 speech acts, including the aircraft identification. All pilots were instructed to read back and execute the ATC instructions. Readback errors and requests to clarify ATC messages were the primary measures of pilot communication. Readback strategies, such as whether pilots repeated instructions in the same format as issued by ATC, were also examined. We found only limited evidence that the grouped format improved pilot memory for ATC messages. In one analysis of requests for clarification, pilots who received grouped instructions produced fewer requests than did pilots who received the same instructions in sequential format, suggesting that they were less likely to misunderstand the ATC messages. Pilots who received grouped instructions were also more likely to read back the grouped instructions in sequential format, suggesting that prior experience with the sequential format influenced pilot communication in this study. ATC message length had a more clear-cut influence on pilot communication, with readback errors and requests for clarification increasing for longer messages.

17. Key Words General Aviation Communication, ATC Communication, Short-Term Memory, Aging	18. Distribution Statement Document is available to the public through the National Technical Information Service, Springfield, Virginia 22161.

19. Security Classif. (of this report) UNCLASSIFIED	20. Security Classif. (of this page) UNCLASSIFIED	21. No. of Pages 27	22. Price

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

ACKNOWLEDGMENTS

We thank Scott Bickford for coding the pilot-controller communication data; and Dennis Beringer and Judith Burki-Cohen for their insightful comments on earlier drafts of this report. We also acknowledge Delbert Marsh, Monte Taets, and Dr. Steve Wreggit for running pilot participants through the protocol, and Sandy Dooty for transcribing the audiotapes.

TABLE OF CONTENTS

IMPROVING PILOT/ATC VOICE COMMUNICATION IN GENERAL AVIATION

"The chief virtue that language can have is clearness, and nothing detracts from it so much as the use of unfamiliar words."
—Hippocrates

1.0 INTRODUCTION

1.1 Background

The present study investigated the influence of format for numerical information (grouped vs. sequential presentation) on General Aviation pilot communication in a simulated flight environment. While infrequent, miscommunication between pilots and controllers is a persistent problem in the National Airspace System. Communication problems arise in part because complex air traffic control (ATC) messages sometimes overload pilot memory. For example, incorrect readbacks tend to increase with message length, in part because longer messages increase the chance of confusion or interference among parts of the message (e.g., heading and speed instructions) in working memory (see Morrow & Rodvold, 1998; Prinzo & Britton, 1993, for reviews).

This study investigated the hypothesis that grouped presentation of numerical information reduces memory load on pilots. This issue was examined because the grouped format is now used by controllers in certain circumstances. FAA Order 7110.65L (Air Traffic Control, Section 4, Radio and Interphone Communications) defines "Group form" as "the pronunciation of a series of numbers as the whole number, or pairs of numbers they represent rather than pronouncing each separate digit," and "The number '0' is pronounced as 'zero' except where it is used in approved 'group form' for authorized aircraft call signs (e.g., EMAIR One Ten), and in stating altitudes (e.g., Ten thousand five hundred)." Paragraph 2-4-17, NUMBER USAGE, states that serial numbers are to be spoken as separate digits and that "Altitudes may be restated in group form for added clarity if the controller chooses." Paragraph 2-4-18, NUMBER CLARIFICATION, states, "If deemed necessary for clarity, and after stating numbers as specified in Para 2-4-17, controllers may restate numbers using either group or single-digit form." However, little is known about the impact of changing this format on ATC communication.

There is abundant evidence from laboratory studies that grouping cues (e.g., pauses, grouped pronunciation of numerical information) allow people to recode information into larger "chunks" in working memory (Crowder, 1976). Chunking is a strategy by which individual items are combined into fewer units called "chunks." For example, the digits 1776149219181941 are virtually impossible to repeat verbatim. When grouped into 4-digits per chunk, the ability to repeat all the digits as chunks improves. However, evidence for benefits of grouping on memory for ATC materials is mixed. Loftus, Dark, and Williams (1979) examined non-pilot memory for ATC clearances of varying lengths, with the four digits making up transponder codes presented either grouped or sequentially. While grouping did not improve memory for the transponder code, it did boost memory for radio frequencies presented in the same message. Loftus et al. (1979) suggested that the two kinds of information were uniquely encoded, which reduced interference in working memory. In Parker-Haney (1991), airline pilots listened to, repeated, and then recalled single instructions after varying retention intervals. All instructions were presented either in grouped or sequential format, using a within-subject design. In the grouped condition, digits were presented in pairs and followed by a short pause. For example, "3163" was presented as "31," pause, "63." There were no pauses when digits were presented sequentially, in a digit-by-digit format such as "3-1-6-3." Grouping did not improve memory in this study.

In a more comprehensive study of the impact of grouped formats on pilot memory, Burki-Cohen (1995) had airline pilots listen to, read back, and enter into a Mode Control Panel, ATC messages that varied in length (3 to 5 instructions) and format (grouped, sequential, restated in both formats) using a within-subject design. Not surprisingly, restated instructions were better remembered. Grouping did

not improve memory, and it even reduced memory for longer messages. This study did find that speed instructions were better remembered when presented in grouped format. Burki-Cohen suggested that this format reduced confusion with heading instructions in the same message, which is consistent with the unique encoding hypothesis proposed by Loftus et al. (1979).

There are several reasons to believe the studies underestimated the impact of the grouping format on pilot memory. First, prior experience with grouped format was not addressed. Pilots may need practice to develop efficient recoding strategies based on grouping cues. Theories of skilled memory emphasize the role of practice in developing recoding or chunking strategies that overcome working memory limits (e.g., Ericsson & Pennington, 1993.) Practice may be especially important for taking advantage of grouped format because pilots are more familiar with the sequential format, which is more common in the current ATC system. It is interesting that the participants in the Loftus et al. study (1979) demonstrated grouping benefits, and they were highly practiced before the experimental trials began. They were also nonpilots, which avoided the possible interfering effects of prior experience with ATC communication. This raises issues related to training both new pilots and experienced pilots to take advantage of a new format for ATC communication (see Discussion).

Second, in both the Parker-Haney and Burki-Cohen studies, most instructions were presented in both grouped and sequential formats (in different messages). It may be important to consistently associate grouping cues with only certain types of instructions (those most likely to be grouped in actual operations, e.g., altitude). This allows unique coding that may minimize interference between parts of messages in working memory, making them less similar. Again, Loftus et al. (1979) restricted grouping to one type of command.

Third, the previous studies did not test pilots under flying conditions. Pilots may need at least a rudimentary flight context to develop and utilize strategies that capitalize on grouping cues that will generalize to actual flying conditions.

Finally, the impact of grouped format should be examined under a range of communication conditions so that the findings will have greater generalizability. For example, grouped format may be especially useful for longer messages because it minimizes the build-up of interference in working memory (Loftus et al., 1979; see also Burki-Cohen, 1995).

The primary objective of the present study was to investigate if grouping numbers in an ATC message improves General Aviation pilot communication under realistic flight conditions. Radio frequency and altitude instructions were selected because they are so prevalent in ATC transmissions, and pilot readbacks of these types of instructions are more likely to contain non-standard communication (Prinzo, 1996) or readback errors (Cardosi, 1996). Radio frequency and altitude instructions were presented in ATC messages, which varied in length from 2 to 5 speech acts (i.e., part of the message with a single function, such as a call sign or instruction; see Morrow, Lee, & Rodvold, 1993; Prinzo, Britton, & Hendrix, 1995), and were presented in grouped format in the Grouping condition only. They were classified as "critical instructions." Heading, altimeter, transponder code, and all other types of instructions were always spoken in sequential format in both the Grouping and Control conditions; they were classified as "other instructions." In this report, the terms "Experimental Condition" and "Instruction Format" will be used interchangeably to improve the readability of this document.

1.2 Hypotheses

We examined the following hypotheses concerning the influence of ATC instruction format, message length, and instruction type (critical or other) on pilot communication performance. Performance was measured by the proportion of readback errors, requests to clarify ATC messages, and several readback strategies.

H1A: Selective Effects of Instruction Format. Grouped format will improve pilot memory only for the critical instructions presented in this format. Pilots should demonstrate better memory for critical but not other instructions within messages in the Grouping vs. Control condition. This will occur if the grouped format enables more efficient memory coding of the altitude and radio frequency instructions (compared with other types of instructions, which were always presented sequentially).

H1B: General Effects of Instruction Format. Grouped format will improve pilot memory for both critical and other instructions (e.g., better memory for

headings as well as altitudes in the Grouping vs. Control condition). This will occur if unique coding of the Critical vs. Other Instructions in the same message reduces interference in working memory (cf. Loftus et al., 1979).

H2: Message Length. Pilots will remember longer ATC messages less accurately than shorter messages, presumably because more information must be retained in working memory, which increases the chances of interference (e.g., Cardosi, 1993; Morrow et al., 1993).

H3: Joint Effects of Instruction Format and Message Length. The grouped format will mitigate decrements associated with message length (i.e., message length effect is smaller for the Grouping rather than Control condition). These results will help clarify the conditions under which the grouped format improves pilot memory for ATC instructions and pilot performance.

2.0 METHOD

2.1 Participants

Twenty-four adults (3 women and 21 men) who were native speakers of English were recruited as paid volunteers. All participants held current pilot medical certificates, and 21 were instrument-rated. The participants had flown a mean of 1311 total flying hours (std. = 1509.0), with a mean of 72.9 hours (std. = 75.0) flown in the 90 days previous to the study. They had flown a mean of 210 IFR hours (std. = 463.0), with a mean of 9.6 hours (std. = 12.0) flown in the 90 days previous to the study. Mean participant age was 28.8 (std. = 8) and mean level of education was 15.1 years (std. = 1.5).

Participants were randomly assigned to the Grouping or Control condition. Appendix A shows that differences between participants in the two conditions were not significant for total overall flying hours, total IFR hours, recent overall flying hours, recent IFR hours, level of education, Weschler Adult Intelligence Scales-Revised (WAIS-R) Forward Digit Span score, or Backward Digit Span score. The digit span test, where participants listen to and repeat sequences of digits (of varying lengths), is a standard measure of short-term memory ability. It was used to check if participants in the two conditions differed in short-term memory ability, which is involved in pilot communication (Taylor, Yesavage, Morrow, Dolhert, Brooks, & Poon, 1994). The digit span test also allowed us to examine whether pilot communication performance in the study was influenced by individual differences in memory capacity. Despite the random assignment procedure, participants in the Grouping condition were older and had flown for more years than participants in the Control condition.

2.2 Basic General Aviation Research Simulator (BGARS)

The BGARS is a medium-fidelity, fixed-base, computer-controlled flight simulator (see Beringer, 1996). The controls and displays used in the BGARS simulated those of a Beechcraft Sundowner. Control inputs were provided by analog controls, including a damped and self-centering yoke, navigation radio frequency selection module, rudder pedals, throttle, flap control, and trim control. Instruments were displayed on a cathode ray tube (CRT) and reacted in real time to all control inputs and aircraft conditions. The external views consisted of a 50-degree forward-projected view and two smaller left-view CRTs. The smaller CRTs provided the pilot with peripheral external views that projected from 25-degrees out to about 112-degrees left of the centerline. The simulation included two-way radio communication that enabled pilots to talk with simulated Tower and Terminal Area controllers. Pilots listened to and read back ATC messages over a standard-issue aviation headset. The earphones and attached microphone are designed to reduce noise reception and transmission.

2.3 ATC Messages

During each flight, the pilot communicated with simulated controllers at local air traffic control towers and the terminal radar approach control (TRACON) facility. For each mission, 29 scripted ATC messages were developed and recorded by an ATC instructor from the FAA Academy. Nineteen of those scripted messages varied in length and contained altitude or radio frequency instructions or both. Short (3 speech acts) and long (5 speech acts) ATC messages containing altitude or radio frequency instructions or both were roughly balanced for frequency of occurrence throughout the mission and were classified as "critical" messages. There were 38 critical messages per subject across the two missions. The content of each message was organized according to the FAA Order 7110.65, Air Traffic Control. The instructor also played the role of a controller during

Table 1. Examples of Radio Frequency and Altitude Instructions Presented in Grouped Format

Symbol Key: C = call sign; W = weather information; R = runway; H = heading; A= altitude SQ = squawk; F = radio frequency

PHASE OF FLIGHT: TAKEOFF	
PWA TWR	Sundowner one niner Two Golf Bravo, wind one niner zero at one two, runway one seven left, cleared for takeoff (3) {C, W, R}
PWA TWR	Sundowner Two Golf Bravo, turn right heading two eight zero, climb and maintain ***forty-one hundred*** maintain VFR, squawk zero three two four, contact Oklahoma City approach ***one twenty-three point sixty five*** (5) {C, H, A, SQ, F} **[1]
PHASE OF FLIGHT: 1st PRACTICE APPROACH	
OKC APCH	Sundowner one niner Two Golf Bravo, IDENT, maintain ***thirty-eight hundred***, information kilo current altimeter two niner niner eight (4) {C, A, Altimeter} *[2]
OKC APCH	Sundowner Two Golf Bravo, radar contact two miles southwest of Wiley Post airport, turn right heading three five zero vector to ILS runway one seven right final approach course (4) {C, H, R}
OKC APCH	Sundowner Two Golf Bravo, descend and maintain ***thirty-five hundred***, contact approach ***one thirty-one point eight*** (3) {C, A, F} **
OKC APCH	Sundowner Two Golf Bravo, how will this approach terminate? (2) {C}
OKC APCH	Sundowner Two Golf Bravo, upon completion of low approach, turn left heading zero eight zero, climb and maintain ***forty-one hundred***, contact approach ***one twenty-six point seven*** (4) {C, H, A, F} **
OKC APCH	Sundowner Two Golf Bravo, turn right heading one zero zero, descend and maintain ***thirty-one hundred*** (3) {C, H, A} *
OKC APCH	Sundowner Two Golf Bravo, turn right heading one four zero, descend and maintain ***twenty-seven hundred*** (3) {C, H, A}
OKC APCH	Sundowner Two Golf Bravo, five miles from outer marker, maintain ***twenty-seven hundred*** until established on the localizer, cleared ILS runway one seven right approach, contact tower ***one thirty-two point two*** (5) {C, A, R, F}
PHASE OF FLIGHT: TWR CLEARANCE	
OKC TWR	Sundowner Two Golf Bravo, wind one six zero at one two, runway one seven right cleared low approach (3) {C, W, R}
OKC TWR	Sundowner Two Golf Bravo, Oklahoma City altimeter two niner niner seven, contact approach ***one twenty-four point forty-five*** (3) {C, Altimeter, F} *
PHASE OF FLIGHT: 2nd PRACTICE APPROACH	
OKC APCH	Sundowner Two Golf Bravo, radar contact (2) {C}

* Presence of either altitude or radio frequency instruction

** Presence of both altitude and radio frequency instruction

[1] The altitude and maintain VFR could be considered as one and the same in the instance of a practice approach, i.e., one speech act. Its purpose was to reinforce to the pilot that IFR-type handling or service was being provided; this would confirm or clarify that IFR separation was not being provided. Per FAR Part 91, GENERAL OPERATING AND FLIGHT RULES, pilot responsibility is still to maintain clear of clouds.

[2] ATIS includes the ICAO alphabet letter associated with current terminal information of which the altimeter is a part.

Table 2. Examples of Radio Frequency and Altitude Instructions Presented in Sequential Format

Symbol Key: C = call sign; W = weather information; R = runway; H = heading; A= altitude SQ = squawk; F = radio frequency

PHASE OF FLIGHT: TAKEOFF	
PWA TWR	Sundowner one niner Two Golf Bravo, wind one niner zero at one two, runway one seven left, cleared for takeoff (3) {C, W, R}
PWA TWR	Sundowner Two Golf Bravo, turn right heading two eight zero, climb and maintain ***four thousand one hundred***, maintain VFR, squawk zero three two four, contact Oklahoma City approach ***one two three point six five*** (5) {C, H, A, SQ, F} **
PHASE OF FLIGHT: 1st PRACTICE APPROACH	
OKC APCH	Sundowner one niner Two Golf Bravo, IDENT, maintain ***three thousand eight hundred***, information kilo current altimeter two niner niner eight (4) {C, A, Altimeter} *
OKC APCH	Sundowner Two Golf Bravo, radar contact two miles southwest of Wiley Post airport, turn right heading three five zero vector to ILS runway one seven right final approach course (4) {C, H, R}
OKC APCH	Sundowner Two Golf Bravo, descend and maintain ***three thousand five hundred***, contact approach ***one three one point eight*** (3) {C, A, F} **
OKC APCH	Sundowner Two Golf Bravo, how will this approach terminate? (2) {C}
OKC APCH	Sundowner Two Golf Bravo, upon completion of low approach, turn left heading zero eight zero, climb and maintain ***four thousand one hundred***, contact approach ***one two six point seven*** (4) {C, H, A, F} **
OKC APCH	Sundowner Two Golf Bravo, turn right heading one zero zero, descend and maintain ***three thousand one hundred*** (3) {C, H, A} *
OKC APCH	Sundowner Two Golf Bravo, turn right heading one four zero, descend and maintain ***two thousand seven hundred*** (3) {C, H, A}
OKC APCH	Sundowner Two Golf Bravo, five miles from outer marker, maintain ***two thousand seven hundred*** until established on the localizer, cleared ILS runway one seven right approach, contact tower ***one three two point two*** (5) {C, A, R, F}
PHASE OF FLIGHT: TWR CLEARANCE	
OKC TWR	Sundowner Two Golf Bravo, wind one six zero at one two, runway one seven right cleared low approach (3) {C, W, R}
OKC TWR	Sundowner Two Golf Bravo, Oklahoma City altimeter two niner niner seven, contact approach ***one two four point four five*** (3) {C, Altimeter, F} *
PHASE OF FLIGHT: 2nd PRACTICE APPROACH	
OKC APCH	Sundowner Two Golf Bravo, radar contact (2) {C}

* Presence of either altitude or radio frequency instruction
** Presence of both altitude and radio frequency instruction

the simulation by responding to any pilot request for information or clarification of the recorded ATC messages.

Some of the messages that participants heard appear in Tables 1 and 2 with asterisks (*) next to them. A single asterisk denotes the presence of either an altitude or radio frequency; a double asterisk indicates that both occur in the message. The number in parentheses refers to the number of speech acts in the message. Pilots in the study were only required to read back instructions with numerical information. That is, they did not read back non-instructional speech acts such as advisories, reports, or acknowledgments. Instructions that pilots read back are indicated by curly brackets ({}) in Tables 1 and 2 (C = call sign, W = weather information, A = altitude, H = heading, R = runway, SQ = squawk, and F= radio frequency).

Corresponding instructions (e.g., change in altitude, radio frequency) in the two missions were changed to reduce the possibility that pilot experience with the ATC messages in the first mission would influence the second mission. All messages were identical in the Grouping and Control conditions, except that the altitude and radio frequency instructions in the critical messages were presented in grouped format in the Grouping condition and in sequential format in the Control condition. Altitudes and radio frequencies were selected to conform to the FAA Order 7110.65, Air Traffic Control. Radio, ATIS, and UNICOM frequencies used in and around the OKC TRACON were not used to avoid the problem of differential familiarity with these frequencies (and thus potential interference effects) among participants. For radio frequencies, "1" always occurred in the first position, "2" and "3" always occurred in the second position, and the numbers "0" through "9" were equally represented in the third position. Values for the heading instructions were selected to conform to the 360 - degree compass settings and the direction of winds at the departure and destination airports. Altitudes were selected to conform to aircraft performance characteristics flying under visual flight rules and airspace requirements.

2.4 Flight Scenarios

Prior to developing the scripts and missions, the FAA Academy instructor flew in and around the Oklahoma City area. Using his notes, the mission depicted in Figure 1 was constructed. As shown in Figure 1, initial takeoff occurred at Wiley Post Airport (PWA) and the pilot made several practice approaches at Will Rogers Airport (KOKC) before returning to and landing at PWA. Generalities in airspace procedures, practice approaches, and General Aviation flight operations were preserved to the extent possible. Altitude, radio frequency, and other locality-specific, numerical information were changed to meet the design requirements of the study. Participants flew two 1-hour missions in and around the Oklahoma City area. The two missions were variations of the same route.

2.5 Procedures

2.5.1 Training Session. Prior to familiarization training, participants completed a demographic questionnaire about their age, sex, education, pilot licenses, and flight experience. They also completed the Forward and Backward Digit Span sub-scales of the (WAIS-R) test. Then they completed two practice tasks. One task familiarized them with the types of ATC messages they would hear during the experiment (participants in the Control condition only heard sequential formats, while participants in the Grouping condition heard both grouped and sequential formats). The other task provided them with instruction and practice flying the BGARS.

2.5.1.1 Familiarization with and Reading Back ATC Messages. To become familiar with the types of messages and the controller's voice, participants first heard a sequence of single instructions that was presented in grouped or sequential format, depending on their experimental condition. For example, ***Sundowner Two GolfBravo, contact tower one twenty-one point one*** contains two speech acts: the address of the receiver (airplane call sign) and an instruction to change radio frequency that was spoken in grouped format. Next, they listened to and read back a sequence of 70 multi-instruction messages that varied in length and format (similar to those in the practice and experimental missions). For example, ***Sundowner Two Golf Bravo, fly heading zero five zero vector to final approach course, descend and maintain forty-six hundred*** contains a heading instruction spoken in sequential format and an altitude instruction spoken in grouped format.

Participants in the Grouping condition heard a sequence of ATC messages that varied in format and length that occurred equally often during each quarter

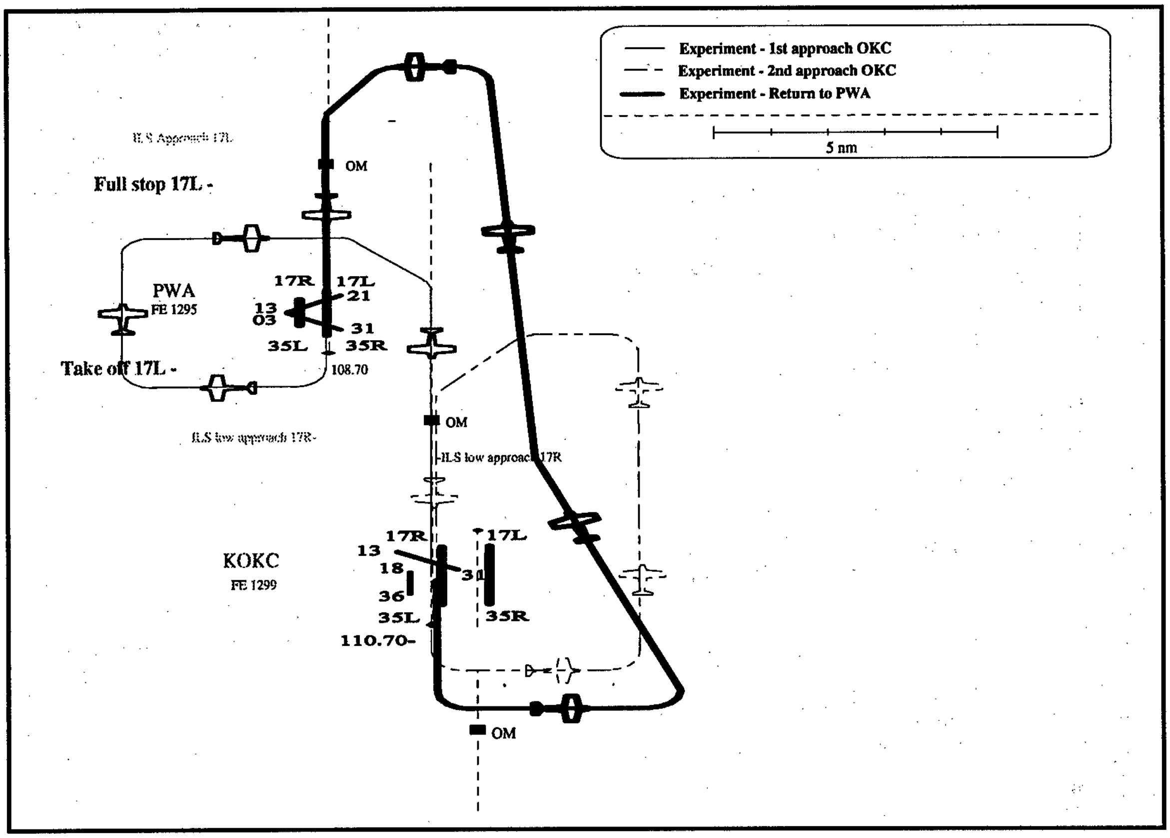

Figure 1. Day 2-Data Collection Flight

of the practice trials. Some of the practice messages in the Grouping condition only contained instructions presented in a sequential format. The same kinds of ATC messages were embedded in a realistic simulator scenario during the practice mission in the BGARS. The same messages occurred in the training session of the Control condition, except the altitude and frequency instructions were presented in sequential format (so that all instructions were presented in sequential format in this condition). In this way, participants in the two conditions received the same amount of training, and the format of the training materials matched the format that they would encounter in the experimental missions. All participants were instructed to read back all instructions in the messages, in both the practice and experimental sessions. This was done to help ensure that everyone used the same criterion for the communication tasks (participants were not told to read back advisories such as weather). Participants spent 20 to 25 minutes listening to and responding to the ATC messages during the training session.

2.5.1.2 Familiarization with the Basic General Aviation Research Simulator. In the second half of the practice session, participants were seated in the BGARS and performed a set of standard maneuvers to become familiar with its performance characteristics. This first encounter with the simulator did not involve communication with ATC. They also flew a 30-minute practice mission that was similar to the missions in the experimental session.

2.5.2 Experimental Session. After the first day of training, participants returned to the laboratory and flew two missions. Before flying, they listened to the messages from the first practice task one more time. All pilot-controller communication during the two experimental missions was recorded on audiotape. In addition, the flight displays were recorded on videotape during the flight, which provided partial information about flight performance. Pilots were not allowed to write down ATC messages during the training and experimental sessions. It was hoped that this requirement would increase the sensitivity of the communication measures, so that different levels of

performance that might be created by the ATC instruction format and message length variables could be detected.

2.6 Experimental Design

The study used a mixed analysis of variance (ANOVA) design with 2 between-group factors and 2 within-subject factors. The between-group factors were Experimental Condition (Grouping or Control) and Mission Sequence (Mission 1 presented First or Second), which was primarily a counterbalancing variable. The two within-subject factors were Instruction Type (Critical or Other) and Message Length (2 to 5 speech acts). The independent and dependent variables presented in Tables 3 and 4 are adapted from Morrow et al. (1993); Morrow, Rodvold, & Lee, (1994a); and Prinzo, Britton, & Hendrix (1995).

2.6.1 Independent Variables. As shown in Table 3, the three major groupings of independent variables pertain to *Subject and Mission, ATC Message, and Speech Act.* If grouped format has a selective benefit on pilot memory for ATC messages, then the responses made by participants in the Grouping condition will be more accurate for critical instructions (i.e., a significant Experimental Condition by Instruction Type interaction). If grouped format has a general benefit, improving pilot memory for all instructions in ATC messages, then the responses made by participants in the Grouping condition will be more accurate for all instructions (i.e., a significant main effect of Experimental Condition).

2.6.2 Dependent Variables. The effect of Experimental Condition (i.e., Instruction Format) and Message Length was examined on several measures of pilot communication. Dependent variables are illustrated in Table 4 with respect to the following sample transaction.

Speaker	**Message**
ATC:	*Sundowner Two GolfBravo/ descend and maintain three thousand five hundred/ contact approach one one eight point seven*
Pilot:	*Down to three thousand five/ contact approach one one eight point seven/ Sundowner*

Table 3. List of Independent Variables

Subject and Mission Variables	**Description**
Instruction Format (Experimental Condition)	Grouping vs. Control condition. In the Grouping condition, altitude and radio frequency instructions were presented in grouped format, and all other speech acts that contained numbers were presented sequentially. In the Control condition, all instructions and other speech acts that contained numbers were presented in sequential format.
Mission Sequence	The presentation order that Mission 1 and Mission 2 were flown.
ATC Message Variables	
Instruction Type	Critical vs. Other instructions. Critical instructions only involved altitude or radio frequency or both aviation topics. Other instructions involved all other aviation topics.
Message Length	The number of speech acts in a message varied from 2 to 5, including call sign.
Speech Act Variables	
Speech Act Type	The type of information in a message was classified by its purpose. Address, Instruction, Request, or Advisory.
Aviation Topic	Speech Act Types were subdivided according to topic (e.g., the topic for an instruction could be altitude, radio frequency).
Speech Act Number	The ordinal position that a speech act occurred in the message.

Table 4. Types of Pilot Communication Measures

Readback Error	ATC Message	Pilot Error
Transposition	3,500	5,300
Intrusion[3]	118.7	135.7
Omission	118.7	118
Substitution	3,500	4,000
Requests for Clarification		
Request Repeat of ATC Message	3,500	Say again (altitude)
Request to Confirm a Readback	3,500	Was that 3,500?
Readback Strategies		
Readback Format Mismatches[4]	Three thousand five hundred	Thirty-five hundred
Incomplete Readbacks[5]	Descend and maintain three thousand five hundred, contact approach 118.7	Down to three thousand / Sundowner/(missing freq.)
Readback Order Mismatches[6]	Descend and maintain three thousand five hundred, contact approach 118.7	Contact approach 118.7 / down to three thousand five hundred

As shown in Table 4, dependent variables were grouped into three primary categories: 1) Readback Errors, 2) Requests for Carification, and 3) Readback Strategies. Pilot readback errors and requests for clarification indicate difficulty with understanding or remembering ATC messages, usually requiring extra radio time to clarify communication (Morrow et al., 1994a). Pilot readback strategies may provide additional information about whether instruction format or message length influenced pilot memory load. Three types of pilot readback strategies were examined: a) Readback format mismatches, b) Incomplete readbacks, and c) Readback order mismatches. Readback format mismatches indicated how often participants changed the format of sequential or grouped ATC instructions in their readbacks. Incomplete readbacks indicated how often pilots read back only part of the message (i.e., they failed to read back at least one instruction). Readback order mismatches indicate how often pilots read back instructions in an order that was different from the order originally presented by the controller. The Aeronautical Information Manual does not require pilots to read back messages in any particular order although individual airlines have their own guidelines and recommended practices for their commercial pilots.

2.7 Coding ATC-Pilot Communication

This section describes the coding and analysis of voice communications between the controller and pilot during each experimental mission.

2.7.1 Coding Procedures. All of the communications between the pilot and controller were transcribed verbatim and then coded by the primary coder, who held a private pilot certificate. The communications were divided into transactions between controller and pilot (defined as a set of messages

[3] Intrusions occur when digits are substituted from a different instruction in the same message. A conservative rule for identifying intrusions was developed: There had to be an overlap of two or more digits between the error and another instruction in the message. If only one digit is confused, the pilot may simply be guessing rather than confusing instructions.

[4] Format of pilot readback mismatches format of ATC instruction.

[5] Readbacks were counted as incomplete if participants failed to repeat instructions; they were not penalized for not repeating advisory information.

[6] Readback order mismatches occur when the read back sequence of speech acts in the message differs from the order in which the instructions were issued by ATC.

between the same controller and pilot with no more than 15 sec of silence between messages). Each transaction was divided into controller and pilot messages (usually a controller clearance or pilot response to the clearance). Then, each message was further divided into speech acts (part of the message with a single function, such as a call sign or instruction; see Morrow et al., 1993; Prinzo et al., 1995). When needed, the primary coder listened to the audiotapes to clarify coding decisions (e.g., intonation may indicate whether a pilot response is a readback or request for clarification).

2.7.2 Inter-Rater Coding Reliability Estimates. One of the principal investigators served as a second coder to establish inter-rater coding reliability. The primary and secondary coders independently parsed and coded the first 18 pilot and ATC messages (52 speech acts) from the first participant's first mission.

2.7.2.1 Parsing Decisions. The two coders agreed on 5 of 6 transaction divisions. The one disagreement occurred when the pilot followed a readback with a request for clarification of the same ATC message, with 35 seconds elapsing from the initial readback to the request. One coder treated the request as part of the same transaction because it referred to the same ATC message as the preceding pilot response. The other coder treated the request as the beginning of a new transaction because of the amount of time that elapsed between the two pilot messages. It was decided that any duration longer than 15 sec between adjacent transmissions would be treated as the beginning of a new transaction. A coding variable was added that indicated to which ATC message each pilot response referred. Percentage agreement between the two coders on dividing messages into speech acts was high (94% agreement based on 52 parsing decisions).

2.7.2.2 Coding. The ATC and pilot messages were coded according to the categories described earlier (see Tables 3 and 4). The lowest agreement was 88% (Instruction format—most disagreements concerned assigning "Not applicable" to ATC advisories). There was 100% agreement for Speech Act Type, 98% agreement for Speech Act Topic, 96% for Readback Format Mismatches, and 100% agreement on Readback Errors and Requests for Clarification.

3.0 RESULTS

3.1 Types of Communication Problems

We first examined the proportion of different types of pilot readback errors and requests for clarification messages. The most frequent type of readback error was Substitution (mean proportion = .04), followed by Omission (.003) and Transposition errors (.002). Of the two types of requests for clarification, Requests to Repeat an ATC message (.08) were more frequent than Requests to Confirm a Readback (.01). Because the different types of pilot readback errors and requests for clarification were so infrequent, we collapsed across subtypes to create a global measure of readback errors and a global measure of requests for clarification. These two global measures were the primary dependent variables in the study. Readback Strategies (Readback Format Mismatch, Incomplete Readback, and Readback Order Mismatch) were also examined.

3.2 Relationships Between Pilot Communication Measures and Individual Difference Variables

To identify possible covariates for the ANOVAs reported below, relationships between the pilot communication measures and the following individual difference variables were examined: participant age, total flight hours, and mean WAIS-R[7] Digit Span. Older participants ($r = .36, p < .05$) and those with more total flight hours ($r = .42, p < .05$) produced more total readback errors. Participant age was positively correlated with total flight hours, ($r = .61, p < .001$) and not with WAIS-R Digit Span ($r = -.15, p = .24$). Participants with higher WAIS-R Digit Span scores produced fewer total requests for clarification ($r = -.45$, $p < .001$) and readback order mismatches ($r = -.32$, $p < .05$), while they produced more readback format mismatches ($r = .47, p < .001$). It is possible that those participants with higher digit span scores had more cognitive resources available to process ATC messages, resulting in fewer understanding problems.

These individual difference variables accounted for a modest percentage of the variance in the communication measures ($R^2 = .10$-$.22$). This finding is not surprising, considering that the study involved a small, heterogeneous sample of pilots, and the individual difference tasks were domain-general while

[7] The WAIS-R Digit Span, as used in this report, is the mean of the participant's forward and backward digit span scores. It was used since these scores were correlated, $r = .35, p < .05$).

the criterion task was aviation-related. Using a sample of pilots with similar characteristics, Taylor et al. (1994) found a similar relationship ($r = -.47$) between WAIS-R Digit Span scores and readback errors during simulated flight.

3.3 Plan of Analysis

Three general types of mixed factor ANOVAs were conducted on the pilot communication measures. All scripted messages were included in the first two analyses to maximize the amount of communication data per participant. First, the effect of Experimental Condition (Grouping or Control), Mission Sequence (Mission 1 presented First or Second), and Instruction Type (Critical or Other) on the communication measures was examined, with Instruction Type, a within-subject factor. The Instruction Type factor compared the critical altitude or radio frequency instructions or both, which were presented in grouped format in the Grouping condition and in sequential format in the Control condition, with all other instructions (e.g., headings and transponder codes). These other instructions were presented in sequential format in both experimental conditions. This first set of analyses, collapsed across Message Length because the levels of the Instruction Type and Message Length factors, was not balanced for all messages (i.e., radio frequency and altitude instructions were not equally likely to occur across messages of different lengths). The second set of analyses included Message Length (2 to 5 speech acts, including call sign) as a within-subject factor and it was collapsed across Instruction Type.

The third set of analyses included only critical messages that contained 3 or 5 speech acts. Because Message Length and Instruction Type were balanced for this subset of messages (frequency or altitude instructions or both occurred equally often in the short and long messages), both factors were included in this analysis. Thus, messages that contained critical instructions were analyzed by mixed-factor ANOVAs with Experimental Condition and Mission Sequence as between-group factors and Instruction Type (Critical vs. Other) and Message Length (3 vs. 5 speech acts) as within-subject factors. Because participants in the Grouping and Control conditions were mismatched on Age (see Appendix A), and both Age and Digit Span scores correlated with several communication measures (but not with each other), Age and Digit Span scores were included as covariates in all ANOVAs. Based on previous research (e.g., Salthouse, 1991), we expected participant Age and Digit Span to have a greater effect on the communication measures under the more difficult message conditions.

3.4 Analyses of Pilot Communication Measures Including ATC Instruction Type

Readback errors, requests for clarification, and readback strategy measures were analyzed by an Experimental Condition (Grouping vs. Control) by Mission Sequence (First vs. Second) by Instruction Type (Critical vs. Other) ANOVA, with the latter a within-subject factor. If consistently presenting a subset of instructions (e.g., radio frequencies and altitudes) in grouped format has a selective effect on pilot communication, then there should be an Experimental Condition by Instruction Type interaction (improving memory only for the grouped instructions embedded within messages). Responses made by participants in the Grouping condition would be more accurate than responses made by participants in the Control condition for instructions spoken in a grouped, but not sequential format. If grouped format improves pilot memory for all instructions (including those that are presented sequentially), there should be a main effect of Experimental Condition, with fewer communication problems occurring in the Grouping rather than Control condition. Table 5 presents mean proportions for the dependent variables by condition for this set of analyses.

3.4.1 Readback Errors. This first analysis did not support either hypothesis about the effect of instruction format on pilot communication performance. The proportion of Readback Errors did not vary with Experimental Condition, $F(1,42) < 1.0$; Mission Sequence, $F(1,42) = 2.2$, $p > .10$; or Instruction Type, $F(1,42) < 1.0$. The Instruction Type by Experimental Condition interaction was not significant, $F(1,42) < 1.0$.[8] As expected by the correlations reported earlier, the Age covariate accounted for a significant amount of variance in Readback Errors, $F(1,42) = 6.7$, $p < .05$, $R^2 = .13$. All other effects were non-significant ($p > .10$).

[8] A test of departure from homogeneity (Mauchly's test of sphericity) was not significant for the readback error measure. Also, studies have shown that the F test is robust to violations of normality and homogeneity of sample distributions (Keppel, 1973).

Table 5. Mean Proportions for Pilot Communication Measures for Analyses with Instruction Type

	Pilot Communication Measures				
			Readback Strategies		
Source	Readback Error	Requests for Clarification	Format Mismatch	Incomplete Readback	Readback Order Mismatch
Experimental Condition					
Grouping	.04	.07	.09	.16	.29
Control	.05	.10	.03	.16	.27
Mission Sequence					
Mission 1 First	.06	.10	.05	.18	.27
Mission 1 Second	.04	.07	.07	.13	.29
Instruction Type					
Critical	.08	.12	.12	--	--
Other	.02	.05	.002	--	--

3.4.2 Requests for Clarification. Requests for clarification did not vary with Experimental Condition, $F(1,42) = 1.9, p > .10$; Mission Sequence, $F(1,42) = 2.4, p > .10$; or Instruction Type $F(1,42) = 2.4, p > .10$. The Instruction Type by Experimental Condition interaction was not significant, $F(1,42) < 1.0$. The Instruction Type by Mission Sequence interaction was significant, $F(1,42) = 7.2, p < .05$. There were more requests to clarify messages that contained critical instructions when Mission 1 was first rather than second (First = .16, Second = .09 mean proportion requests for clarification), but this was not the case for messages that contained other instructions (First mission = .04, Second = .05 mean proportion requests for clarification). Finally, the Digit Span covariate had a significant effect, $F(1,42) = 7.5, p < .01, R^2 = .20$). As the correlations suggest, participants with higher digit span scores requested fewer clarifications of the ATC messages.

3.4.3 Readback Strategies. The effect of instruction format on the proportion of readback format mismatches was examined. The main effect of Experimental Condition approached significance, $F(1,42) = 3.9, p < .06$. Table 5 shows a tendency for pilots in the Grouping condition to change instruction formats more often than pilots in the Control condition. While Instruction Type did not influence mismatches, $F(1,42) < 1.0$, the Instruction Type by Experimental Condition interaction approached significance, $F(1,42) = 3.6, p < .10$.[9] This interaction suggests that, when reading back messages, participants in the Grouping rather than Control condition were more likely to change instruction formats in their readback. They primarily changed the format of critical instructions that were presented in grouped format in the Grouping condition (Grouping = -.17, Control = .07 mismatches). The format of other instructions was rarely changed in either condition (Grouping = .003, Control = .00 mismatches). In other words, participants in the Grouping condition tended to read back sequentially those altitude or radio frequency instructions or both when originally presented in grouped format. One interpretation of this pattern of findings is that participants who heard the relatively novel grouped format tended to change the format back to the traditional sequential form in the readbacks (despite the training they received on the grouped format before the experimental session). This suggests an effect of prior experience with the sequential format.

[9] Error terms for comparisons and simple main effect analyses were based only on the observations in the subset of data being tested, rather than using a pooled error term, as recommended by Keppel (1973).

There was also an effect of Digit Span, $F(1,42) = 10.9$, $p < .01$, $R^2 = .22$; participants with higher digit span scores were more likely to change the instruction format. It may be that changing the instruction format during readbacks requires additional cognitive resources and those participants with more resources (as indexed by digit span scores) were better able to transform the information.

The analyses of incomplete readbacks and readback order mismatches did not include Instruction Type because this factor applied to individual instructions (varying within messages), while the incomplete readback and readback order mismatches applied to the total message. Incomplete readbacks did not vary with Experimental Condition, $F(1,42) < 1.0$, or Mission Sequence, $F(1,42) = 1.3$, $p > .10$.

Readback order mismatches also did not vary with Experimental Condition or Mission Sequence, $F(1,42) < 1.0$ for both factors. However, the Digit Span covariate was significant, $F(1,42) = 4.1$, $p < .05$, $R^2 = .10$. The correlations reported earlier show that participants with higher digit span scores were less likely to change the presented order of instructions while reading back instructions.

3.5 Analyses of Pilot Communication Measures Including ATC Message Length

The results in this section pertain to the effect of Experimental Condition, Mission Sequence, and Message Length on the pilot communication measures (collapsing across Instruction Type). In addition to examining the overall effect of instruction format on communication (main effect of Experimental Condition), we examined if longer messages lead to more pilot communication problems. Table 6 presents the mean proportions for the dependent variables by condition for this set of analyses.

3.5.1 Readback Errors. Readback Errors did not vary with Experimental Condition, $F(1,42) < 1.0$, or with Mission Sequence, $F(1,42) = 2.0$, $p > .10$. The Age covariate accounted for a significant amount of variance, $F(1,42) = 7.7$, $p < .01$.

Message Length did influence Readback Errors, with errors increasing for longer messages, $F(3,126) = 3.2$, $p < .05$. A significant quadratic trend component, $F(1,42) = 4.6$, $p < .05$, reflected that the proportion of Readback Errors tended to level off for messages with 4 to 5 speech acts. The effect of Message Length also depended on participant Age, $F(3,126) = 3.2$, $p < .05$. This interaction was analyzed by comparing readback errors for participants above and below the median age (26 years) for each message length (collapsed across the two experimental conditions). Older participants made significantly more readback errors for messages with only three speech acts, $t(19) = 3.0$, $p < .01$ (see Table 7).

3.5.2 Requests for Clarification. Requests for clarification did not vary with Experimental Condition, $F(1,42) = 1.7$, $p > .10$, or Mission Sequence, $F(1,42) = 1.9$, $p > .10$. The Digit Span covariate accounted for a significant amount of variance, $F(1,42) = 8.4$, $p < .01$. Requests for clarification increased with Message Length, $F(3,126) = 3.9$, $p < .05$ (quadratic trend, $F(1,42) = 8.6$, $p < .01$). The Experimental Condition by Message Length interaction approached significance, $F(3,126) = 2.5$, $p < .07$. Because we had predicted that the effect of instruction format would depend on message length, planned comparisons were conducted, although the interaction was not significant. As shown in Table 8, differences in the mean proportion of requests for clarification between the Grouping and Control conditions were not significant for any individual Message Length condition, all $t(23) < 1.3$, $p > .10$.

The effect of Message Length on requests for clarification also depended on Age, $F(3,126) = 4.3$, $p < .01$. As with Readback Errors, this interaction was analyzed by comparing the proportion of requests for clarification by participants above and below the median age. As shown in Table 9, older participants made more requests than younger participants for the longest messages, $t(19) = 2.3$, $p < .05$ but not for the shorter messages, $t(19) < 1.0$ for all comparisons.

3.5.3 Readback Strategies. Participants in the Grouping condition were more likely than those in the Control condition to change instruction formats in their readbacks, $F(1,42) = 4.3$, $p < .05$. The analysis, including Instruction Type described in Section 3.4.3, suggested that participants in the Grouping condition were likely to read back messages that contained grouped ATC instructions in sequential format. There was also an effect of digit span scores in the present analysis, $F(1,42) = 9.6$, $p < .01$.

Readback format mismatches also depended on Message Length, $F(3,126) = 2.9$, $p < .05$, with a trend for mismatches to increase with the length of the messages, linear trend $F(1,42) = 5.4$, $p < .05$. This influence of Message Length also depended on Digit Span, $F(3,126) = 6.3$, $p < .001$. This Message Length by Digit Span interaction was analyzed by comparing

Table 6. Mean Proportions for Pilot Communication Measures for Analyses with Message Length

	Pilot Communication Measures				
			Readback Strategies		
Source	Readback Error	Requests for Clarification	Format Mismatch	Incomplete Readback	Readback Order Mismatch
Experimental Condition					
Grouping	.03	.06	.07	.12	.34
Control	.04	.08	.02	.15	.30
Mission Sequence					
Mission 1 First	.05	.08	.04	.16	.31
Mission 1 Second	.03	.06	.05	.11	.33
Message Length					
2 Speech Acts	.00	.01	.01	.01	--
3 Speech Acts	.03	.03	.07	.03	.15
4 Speech Acts	.06	.08	.07	.11	.24
5 Speech Acts	.07	.17	.04	.39	.56

Table 7. Mean Proportion Readback Errors Presented by Participant Age and Message Length

	Message Length		
Source	3 Speech Acts	4 Speech Acts	5 Speech Acts
Participant Age			
Older participants (N = 10, Age = 36.5)	.10	.08	.07
Younger participants (N = 10, Age = 22.5)	.04	.04	.07

Table 8. Mean Proportion Requests for Clarification Presented by Experimental Condition and Message Length

	Message Length			
Source	2 Speech Acts	3 Speech Acts	4 Speech Acts	5 Speech Acts
Experimental Condition				
Grouping	.02	.02	.06	.14
Control	.00	.03	.10	.19

Table 9. Mean Proportion Requests for Clarification Presented by Participant Age and Message Length

	Message Length			
Source	**2 Speech Acts**	**3 Speech Acts**	**4 Speech Acts**	**5 Speech Acts**
Participant Age				
Older participants	.03	.03	.10	.22
Younger participants	.00	.04	.07	.12

readback format mismatches for participants above and below the median digit span score (median span = 8.0). High and low digit span groups did not differ in age, education, or flight hours. Although the results for the longer messages suggested that participants with higher digit span scores produced more readback format mismatches than participants with lower digit span scores, the differences were not significant for any of the message lengths, $t(19) < 1.6$, $p > .10$ for all comparisons.

The number of incomplete readbacks did not vary with Experimental Condition, $F(1,42) < 1.0$, or Message Length, $F(3,126) < 1.0$. Messages with 2 speech acts (i.e., call sign plus 1 instruction) and incomplete readbacks with only 1 instruction were excluded from the analysis of Readback Order Mismatches because an order analysis requires at least 2 instructions in the readback. Readback order mismatches did not vary with Experimental Condition, $F(1,42) < 1.0$, or Mission Sequence, $F(1,42) < 1.0$. Readback order mismatches did, however, vary with Digit Span scores, $F(1,42) = 4.6$, $p < .05$. Participants were also more likely to read back instructions in a different order for longer ATC messages; linear trend component $F(2,84) = 5.7$, $p < .05$. It is possible that the effect of message length on readback order mismatch reflects a differential recency effect. However, this possibility cannot be assessed in the present study because readback order mismatches were coded as a 0/1 variable, rather than coding the actual order in which instructions were read back when a mismatch occurred.

3.6 Analyses of Critical Messages Including Both Instruction Type and Message Length

The final set of analyses examined only long (5 speech acts) and short (3 speech acts) messages with equal numbers of critical instructions (altitude or radio frequency presented or both in grouped format in the Grouping condition and sequentially in the Control condition). Therefore, Message Length and Instruction Type were balanced for these analyses. The pilot communication measures were analyzed by a mixed-factor ANOVA with Experimental Condition and Mission Sequence as between-group factors and Instruction Type (Critical or Other) and Message Length (Short or Long) as within-subject factors. Participant Age and Digit Span scores were included as covariates. Table 10 presents the mean proportions for the dependent variables by condition for this set of analyses.

In addition to examining whether the effect of instruction format had a selective or a general effect on communication and whether message length increased communication problems, these analyses allowed us to test whether instruction format effects depended on message length. This subset of data was more controlled than the total data set in the sense that critical instructions occurred equally often in the longer and shorter messages. Therefore, error rates may be higher because previous research has demonstrated that communication problems are especially likely for these instructions (Cardosi, 1996; Prinzo, 1996). Thus, the present analyses may be more sensitive to effects of the independent variables. On the other hand, this analysis reduces the chances of finding an effect of message length because the range of this variable was truncated.

3.6.1 Readback Errors. Readback Errors did not vary with Experimental Condition, $F(1,42) < 1.0$, Mission Sequence, $F(1,42) = 2.3$, $p > .10$; Instruction Type, $F(1,42) < 1.0$; or Message Length, $F(1,42) < 1.0$. The Experimental Condition by Instruction Type and the Experimental Condition by Message Length interactions also were not significant, $F(1,42) < 1.0$ for both.

Table 10. Mean Proportions for Pilot Communication Measures for Analyses with both Experimental Condition and Message Length

	Pilot Communication Measures		
Source	**Readback Error**	**Requests for Clarification**	**Readback Format Mismatch**
Experimental Condition			
Grouping	.05	.07	.08
Control	.04	.13	.03
Mission Sequence			
Mission 1 First	.05	.10	.06
Mission 1 Second	.04	.09	.05
Instruction Type			
Critical	.06	.12	.11
Other	.03	.08	.05
Message Length			
3 Speech Acts	.03	.03	.04
5 Speech Acts	.06	.16	.07

3.6.2 Requests for Clarification. This analysis provided the only evidence that pilot comprehension and memory for ATC messages was improved by presenting instruction speech acts in a grouped format. Participants in the Grouping condition made fewer requests for clarification than participants in the Control condition, $F(1,42) = 4.1$, $p < .05$. This advantage of the grouped format was not limited to instructions that were presented in grouped format because the Experimental Condition by Instruction Type interaction was not significant, $F(1,42) < 1.0$. The effect of the Age covariate approached significance, $F(1,42) = 3.7$, $p < .07$.

Neither the main effect of Experimental Condition, $F(1,42) < 1.0$, nor Message Length were significant, $F(1,42) = 1.0$. However, there was a significant Experimental Condition by Message Length interaction, $F(1,42) = 5.6$, $p < .05$ (see Table 11). The effect of instruction format was greater for the longer messages than for the shorter messages. The same pattern occurred for the analysis with all 5 levels of the Message Length variable, although that interaction was not significant. This pattern of message length effects suggests that the influence of grouped format on requests for clarification was more apparent for the longer messages. The benefits of differential coding of the information may be more likely to occur for longer messages, which impose additional demands on working memory.

The effect of Message Length also depended on Age, $F(1,42) = 6.8$, $p < .05$. To analyze this interaction, the proportion of requests for clarification made by older and younger participants after short and long messages was compared. Older participants produced more requests for clarification than younger participants for longer messages (Older = .21, Younger = .12 mean proportion requests for clarification), $t(19) = 2.4$, $p < .05$, but not for shorter messages (Older = .04, Younger = .03 mean proportion requests for clarification), $t(19) < 1.0$.

3.6.3 Readback Format Mismatches. As shown by the aggregate means presented under the title "Marginals" in the final column of Table 12, there were more readback format mismatches under the Grouping condition than under the Control condition, $F(1,42) = 4.5$, $p < .05$. A marginally significant Experimental Group by Instruction Type interaction, $F(1,42) = 3.6$, $p < .07$, suggested that

Table 11. Mean Proportion Requests for Clarification Presented by Experimental Condition and Message Length

Source	Critical Message Length		Mean Experimental Condition
	3 Speech Acts	5 Speech Acts	
Experimental Condition			
Grouping	.02	.14	.07
Control	.04	.19	.13
Mean Message Length	.03	.16	

Table 12. Mean Proportion Readback Format Mismatches Presented by Experimental Condition and Message Length

Source	Message Length		Marginals
	Short	Long	
Grouping Condition	.11	.06	.08
Critical	.20	.10	.15
Other	.01	.01	.01
Control Condition	.04	.03	.03
Critical	.08	.06	.07
Other	.00	.00	.00

participants in the Grouping condition were more likely than participants in the Control condition to change the format of altitude and frequency instructions, which were presented in grouped format (Grouping = .15, Control = .07 readback format mismatches). Participants in both conditions rarely changed the format of Other instructions, which were always presented sequentially (Grouping = .01, Control = .00 readback format mismatch).

A significant Experimental Group by Message Length interaction, $F(1,42) = 5.2$, $p < .05$ (see the 4 means presented in the shaded areas in Table 12) suggested that participants in the Grouping condition were more likely to change the format of instructions in shorter, rather than in longer messages. Participants in the Control condition rarely changed speech act format for either short or long messages. Finally, an Experimental Group by Instruction Type by Message Length interaction, $F(1,42) = 6.5$, $p < .05$, showed that the effect of Message Length in the Grouping condition was restricted to altitude and radio frequency instructions that were spoken in a grouped format. Participants in the Control condition who heard all of the altitude and radio frequency instructions spoken sequentially rarely changed speech act format, regardless of Instruction Type or Message Length. There was also an effect of digit span scores on readback format mismatches, $F(1,42) = 11.6$, $p < .001$. Findings for the readback order and incomplete readback analyses are not reported because they were similar to those from the analyses that used only instruction format or message length.

4.0 DISCUSSION

4.1 ATC Instruction Format

This study provided little support for the primary hypothesis that General Aviation pilots would better remember ATC messages in which some instructions were consistently presented in grouped rather than sequential format. The only supporting evidence comes from the analysis of requests for clarification of critical instructions that contained either 3 or 5 speech acts (including call sign). Participants who received instructions to change altitude or radio frequency or both in grouped format produced fewer requests than participants who received the numbers spoken sequentially as discrete digits. The absence of a significant Instruction Type by Experimental Condition interaction suggests that the grouped format appeared to have a general effect on pilot memory for ATC messages. Perhaps consistently presenting only a subset of instructions in grouped format (i.e., altitudes and frequencies) improves memory for all information in the message. It may be that the unique coding of the altitude and frequency vs. other instructions in the same message reduces interference in working memory (also see Burki-Cohen, 1995; Loftus et al., 1979). The Experimental Condition by Message Length interaction for requests to clarify critical messages suggested that the grouped format might mitigate the effects of message length on pilot memory (see the next section), perhaps by reducing interference effects among parts of the longer messages in working memory. Burki-Cohen (1995), on the other hand, found some evidence that grouped format was especially likely to reduce pilot memory for longer ATC messages. Clearly, more research will be needed to clarify the conditions under which the grouped format helps or hinders pilot communication.

The readback format mismatch measure clearly showed that participants in the Grouping and the Control conditions processed ATC messages differently. While the grouped format had a general effect on requests for clarification, this format appeared to have a more selective effect on the strategy of changing the ATC instruction format when reading back ATC messages. Participants in the Grouping condition primarily changed the format of instructions presented in grouped rather than sequential format, which suggests that past experience with sequential formats influenced how pilots processed the ATC messages. Despite having received training on grouped formats, participants translated the relatively novel format back into the sequential format typically used in ATC communication. The fact that participants with higher digit span scores were more likely than those with lower digit span scores to change formats suggests that this translation process required additional cognitive resources, which could have interfered with message comprehension (however, correlations between readback format mismatches and readback errors or requests for clarification were not significant in the present study). Additional training on listening and reading back grouped instructions might eliminate this translation process and might increase the benefits of grouped format on pilot comprehension and memory for ATC messages.

Before making conclusions about the effect of grouped format on pilot memory, the reader should remember that significant differences between the Grouping and Control conditions occurred only for those messages that contained 3 or 5 speech acts. The benefits of grouping were not realized for the full set of 28 scripted messages (i.e., other messages and messages that contained 2 or 4 speech acts). Differences also were not found for readback errors, perhaps because there were so few in the present study. The practice data, which produced higher readback error rates than in the experimental missions, may yield further evidence for benefits of ATC instruction format on pilot communication; those data will be analyzed.

Even the limited support for benefits of grouped format on pilot memory contrasts with previous studies, which found no evidence that grouped format improved memory (Burki-Cohen, 1995; Parker-Haney, 1991). Several differences in the procedure and design of the present study may explain this difference.

First, participants in the present study practiced responding to ATC messages before the experimental flights. They were more likely to develop recoding strategies based on the grouping cues (although this practice appeared to be insufficient to overcome effects of past experience).

Second, by restricting the grouped format to a subset of instructions, this study encouraged unique encoding of information in the ATC messages that should reduce the build up of interference in working memory. Indeed, the Loftus et al. study (1979), which found some evidence for grouping benefits, provided both high levels of practice and also restricted

grouping to one type of ATC instruction. It is also interesting that the Loftus study involved non-pilots, which eliminated the possibility that prior experience with sequential formats would create interference with the less familiar grouped format.

Third, by examining performance in the context of a flight, the present study may have encouraged pilots to develop strategies that could be used during actual flight conditions. However, it is possible that the flight context might discourage the use of grouping cues because the constraints associated with actual flying conditions might facilitate communication to such an extent that pilots would not need to rely on these cues. For example, some General Aviation and Commercial pilots will write down parts of ATC messages as they are being given by controllers. Also, Jeppesen charts include the radio frequencies and approach and departure procedures for airports. We did not assess these possibilities because the scenarios used in the present study did not involve the procedures actually used at the airports, and we did not measure the participants' familiarity with these procedures.

4.2 ATC Message Length

Like several earlier laboratory studies (Morrow, Rodvold, McGann, & Mackintosh, 1994b) and field studies (e.g., Cardosi, 1993; Morrow et al., 1993), the present study found that communication problems (both readback errors and requests) were more likely to occur for longer ATC messages. These findings suggest that longer messages increased demands on pilot memory. Participants were also likely to re-order instructions in longer messages, presumably as a way of coping with increased demands on working memory (also see Morrow et al., 1994b). The finding that Requests for Clarification correlated with WAIS-R Digit span scores, and that longer messages involved greater age differences for requests, provides further evidence that the ATC messages in this study imposed demands on limited working memory capacity (see below).

4.3 Individual Differences in Communication Performance

Participants with higher WAIS-R Digit Span scores tended to produce fewer requests for clarification and were more likely to change the format of instructions in the ATC messages. Similarly, Taylor et al. (1994) found that General Aviation pilots with higher WAIS-R Digit Span scores made fewer readback errors. It is possible that those participants with higher digit span scores had more cognitive resources available to process ATC messages, and thus were better able to understand and remember the messages (as indexed by requests for clarification). They also had more resources available to reprocess the information, for example, by changing the format of the instructions in the message. Participants with higher digit span scores also were less likely to change the order of instructions in their readback. If re-ordering information (i.e., repeating the last instruction first) is considered a strategy for reducing memory load (Morrow et al., 1994b), pilots with higher digit span scores may have less need to use this strategy.

Age differences were related to both readback errors and requests for clarification. Although it is somewhat surprising to find age differences among pilots within such a young sample (19 to 46 years), it is not unprecedented. Morrow, Leirer, and Yesavage (1990) found that General Aviation pilots primarily in their forties produced more readback errors than pilots in their twenties in a simulated flight environment. Moreover, in the present study, the Age by Message Length interaction for Requests for Clarification is consistent with the Age by Task Complexity interaction predicted by theories that posit age declines in cognitive resources such as processing speed or working memory capacity (Salthouse, 1991).

5.0 CONCLUSIONS

The present study, like several earlier studies, found that longer ATC messages imposed heavier demands on pilot memory and may increase the likelihood of communication problems. Also, there was only limited evidence that grouped instruction format improves pilot memory for ATC messages. However, the benefits of the grouped instruction format may be most evident when demands are high. It is possible that the manner in which digits are spoken, either sequentially or in grouped format, produces subtle effects on pilot comprehension and memory, and would require a more sensitive experimental design to produce more definitive evidence for these effects. For example, a larger sample of participants or more communication data per participant by phase of flight might have produced a clearer picture of the effects of ATC speech act format on pilot performance.

We did find that pilots tended to translate grouped formats into the more familiar sequential format when reading back ATC messages. This finding highlights issues related to introducing a new procedure into an existing environment. It suggests the importance of communication training for overcoming interfering effects of past experience with ATC communication. It also raises the possibility that grouped formats in ATC messages would provide greater benefits for new pilots, who would not have prior experience with sequential formats in ATC communication. Of course, any change in format, or other aspects of ATC communication, would pose a transition problem for pilots trained under the existing system. While participants in the present study were General Aviation pilots with only limited flying experience, the two earlier studies that found little or no benefits of grouping format used highly practiced airline pilots as participants (Burki-Cohen, 1995; Parker-Haney, 1991). Airline pilots might encounter greater carry-over from past experience listening to and reading back numbers in a sequential format and may require more communication training to benefit from grouped format. It is also possible that a change from sequential to grouped formats would reduce the distinctiveness of some numbers (perhaps due to novelty), which could increase the possibility of some confusion in pilot-controller communication. This possibility could not be addressed in the present study because of the limited number of readback errors. Clearly, we need to better understand the potential benefits and costs of using the grouped format for ATC instructions, and the conditions under which these benefits might best be realized, before recommendations for changing ATC instruction format are made.

6.0 REFERENCES

Airman's Information Manual, Official Guide to Basic Flight Information and ATC Procedures. *(1994). Washington, DC: U.S. Government Printing Office.*

Beringer, D. (1996). *Use of off-the-shelf PC-based flight simulators for aviation human factors research.* Federal Aviation Administration, Office of Aviation Medicine Technical Report DOT/FAA/AM-96/15, Washington, DC. Available from: National Technical Information Service, Springfield, VA 22161; ordering no. ADA304890.

Burki-Cohen, J. (1995). *Say Again? How complexity and format of Air Traffic Control instructions affect pilot recall.* Paper presented at the 40th Annual Air Traffic Control Association.

Cardosi, K.M. (1993). *An analysis of en route controller-pilot voice communications.* US Department of Transportation, Office of Research and Development Report DOT/FAA/RD-93/11, Washington, DC.

Cardosi, K.M. (1996). *An analysis of TRACON (Terminal Radar Approach Control) controller-pilot voice communications.* US Department of Transportation, Office of Research and Development Report DOT/FAA/AR-96/66). Washington, DC.

Crowder, R. (1976). *Principles of learning and memory.* Hillsdale, NJ: Erlbaum.

Ericsson, K.A. & Pennington, N. (1993). Experts and expertise. In G. Davis & R. Logie (Eds.), *Memory in everyday life* (pp. 241-72). Amsterdam: North-Holland.

Keppel, G. (1973) *Design and analysis* (pp 75-6). Englewood Cliffs, NJ: Prentice-Hall, Inc.

Loftus, G., Dark, V., & Williams, D. (1979). Short-term memory factors in ground controller/pilot communication. *Human Factors, 21,* 169-81.

Morrow, D., Lee A., & Rodvold, M. (1993). Analyzing problems in routine controller-pilot communication. *International Journal of Aviation Psychology, 3,* 285-302.

Morrow, D.G., Leirer, V.O., & Yesavage, J. (1990). The effect of alcohol and aging on communication during flight. *Aviation, Space, and Environmental Medicine, 61,* 12-20.

Morrow, D. & Rodvold, M. (1998). Communication issues in Air Traffic Control (pp. 421-56). In M. Smolensky & E. Stein (Eds.), *Human Factors in Air Traffic Control.* New York: Academic Press.

Morrow, D., Rodvold, M., & Lee, A. (1994a). Nonroutine transactions in controller-pilot communication. *Discourse Processes, 17,* 235-58.

Morrow, D., Rodvold, M., McGann, A., & Mackintosh, M. (1994b). Collaborative strategies in air-round communication. *Proceedings of Aerotech 94 Conference,* pp.119-24, Technical Paper #942138, Society of Automotive Engineers, Los Angeles, CA.

Parker-Haney, E. (1991). The effects of unique encoding on the recall of numeric information. In R.S. Jensen (Ed.), *Proceedings of the Sixth International Symposium on Aviation Psychology,* Columbus, OH.

Prinzo, O.V. (1996). *An analysis of approach control/pilot voice communications.* Federal Aviation Administration, Office of Aviation Medicine Technical Report DOT/FAA/AM-96/26, Washington, DC. Available from: National Technical Information Service, Springfield, VA 22161; ordering no. ADA274457.

Prinzo, O.V. & Britton, T.W. (1993). *ATC/pilot voice communications: A survey of the literature.* Federal Aviation Administration, Office of Aviation Medicine Technical Report DOT/FAA/AM-93/20, Washington, DC. Available from: National Technical Information Service, Springfield, VA 22161; ordering no. ADA317528.

Prinzo, O.V., Britton, T.W., & Hendrix, A.M. (1995). Development of a coding form for approach control/pilot voice communications. In B.J. Kanki & O.V. Prinzo (Eds.) *Methods and metrics of voice communications.* Federal Aviation Administration, Office of Aviation Medicine Technical Report DOT/FAA/AM-96/10, Washington, DC. Available from: National Technical Information Service, Springfield, VA 22161; ordering no. ADA307148.

Salthouse, T. (1991). Mediation of adult age differences in cognition by reductions in working memory and speed of processing. *Psychological Science, 2,* 179-83.

Taylor, J., Yesavage, J., Morrow, D., Dolhert, N., Brooks, J., & Poon, L. (1994). The effects of information load and speech rate on young and older aircraft pilots' ability to execute simulated Air Traffic controller instructions. *Journal of Gerontology: Psychological Sciences. 49,* 191-200.

APPENDIX A

Demographic Data for Participants

	Experimental Condition						
	Grouped			Control			t-test
Variable	**Mean**	**Range**	**Std**	**Mean**	**Range**	**Std**	**(df = 22)**
Total overall flying hours	1766.6	165-5000	1920.0	855.8	855.8	789.2	1.5
Total IFR hours	321.5	.05-2300	646.8	98.5	98.5	55.8	1.2
Recent overall hours	52.5	0-150	48.4	93.3	93.3	92.4	1.4
Recent IFR hours	8.5	0-27	11.0	10.7	10.7	12.9	<1.0
Years of flying	10.3	1-27	7.7	3.9	3.9	2.6	2.7*
Education (years)	14.8	12-16	1.6	15.4	15.4	1.4	<1.0
WAIS-R Digit Span Score							
Forward	9.2	9.2	2.5	8.8	8.8	2.6	<1.0
Backward	7.5	7.5	2.2	7.1	7.1	2.0	<1.0
Average Digit Span	8.3	8.3	2.2	7.9	7.9	1.5	<1.0
Age	33.0	33.0	9.2	24.5	24.5	3.3	3.0**

* $p < .05$; ** $p < .01$.

☆U.S. GOVERNMENT PRINTING OFFICE: 1999 - 760-082/00004